REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
03-10-2005	Journal Article (refereed)	

4. TITLE AND SUBTITLE
Evaluations of Global Wave Prediction at FNMOC

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
0602435N

6. AUTHOR(S)
Rogers, Erick W., Wittman, Paul, Wang, David, Clancy, R.M., Hsu, Y. Larry

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
73-M123-03-5

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
Oceanography Division
Stennis Space Center, MS 39529-5004

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/JA/7320--04-4

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5660

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
It is a major challenge to determine whether bias in operational global wave predictions is predominately due to the wave model itself (internal error) or due to errors in wind forcing (an external error). Another challenge is to characterize bias attributable to errors in wave model physics (e.g., input, dissipation, and nonlinear transfer). In this study, hindcasts and an evaluation methodology are constructed to address these challenges. The bias of the wave predictions is evaluated with consideration of the bias of four different wind forcing fields [two of which are supplemented with the NASA Quick Scatterometer (QuikSCAT) measurements]. It is found that the accuracy of the Fleet Numerical Meteorology and Oceanography Center's operational global wind forcing has improved to the point where it is unlikely to be the primary source of error in the center's global wave model (WAVEWATCH-III). The hindcast comparisons are specifically designed to minimize systematic errors from numerics and resolution. From these hindcasts, insight into the physics-related bias in the global wave model is possible: comparison to in situ wave data suggests an overall positive bias at northeast Pacific locations and an overall negative bias at northwest Atlantic locations. Comparison of frequency bands indicates a tendency by the model physics to overpredict energy at higher frequencies and underpredict energy at lower frequencies.

15. SUBJECT TERMS
operational global wave predictions, dissipation, hindcast, wind forcing fields, frequencies

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UL	16	W. Erick Rogers
					19b. TELEPHONE NUMBER *(Include area code)* 228-688-4727

Evaluations of Global Wave Prediction at the Fleet Numerical Meteorology and Oceanography Center*

W. ERICK ROGERS
Oceanography Division, Naval Research Laboratory, Stennis Space Center, Mississippi

PAUL A. WITTMANN
Fleet Numerical Meteorology and Oceanography Center, Monterey, California

DAVID W. C. WANG
Oceanography Division, Naval Research Laboratory, Stennis Space Center, Mississippi

R. MICHAEL CLANCY
Fleet Numerical Meteorology and Oceanography Center, Monterey, California

Y. LARRY HSU
Oceanography Division, Naval Research Laboratory, Stennis Space Center, Mississippi

(Manuscript received 3 September 2004, in final form 25 April 2005)

ABSTRACT

It is a major challenge to determine whether bias in operational global wave predictions is predominately due to the wave model itself (internal error) or due to errors in wind forcing (an external error). Another challenge is to characterize bias attributable to errors in wave model physics (e.g., input, dissipation, and nonlinear transfer). In this study, hindcasts and an evaluation methodology are constructed to address these challenges. The bias of the wave predictions is evaluated with consideration of the bias of four different wind forcing fields [two of which are supplemented with the NASA Quick Scatterometer (QuikSCAT) measurements]. It is found that the accuracy of the Fleet Numerical Meteorology and Oceanography Center's operational global wind forcing has improved to the point where it is unlikely to be the primary source of error in the center's global wave model (WAVEWATCH-III). The hindcast comparisons are specifically designed to minimize systematic errors from numerics and resolution. From these hindcasts, insight into the physics-related bias in the global wave model is possible: comparison to in situ wave data suggests an overall positive bias at northeast Pacific locations and an overall negative bias at northwest Atlantic locations. Comparison of frequency bands indicates a tendency by the model physics to overpredict energy at higher frequencies and underpredict energy at lower frequencies.

1. Introduction

Accurate nowcasting and forecasting of ocean wave conditions is one of the primary missions of the U.S. Naval Meteorology and Oceanography Command (CNMOC). Smaller-scale wave models receive boundary conditions from the global model, making the accuracy of the latter particularly essential. Some noteworthy advancement in wave modeling has occurred during the past decade, but most validations suggest that substantial errors (e.g., 40–60-cm root-mean-square error) are typical (see, e.g., Bidlot et al. 2002).

At the present time, there are two wave models being run operationally at global and regional scales by the U.S. Navy: Wave Model (WAM) cycle 4 (e.g., WAMDI Group 1988; Günther et al. 1992; Komen et al. 1994; henceforth denoted WAM4) at the Naval Oceanographic Office (NAVO) and WAVEWATCH III (e.g., Tolman 1991; Tolman and Chalikov 1996; Tolman 2002a, henceforth denoted WW3) at the Fleet Numerical Meteorology and Oceanography Center (FNMOC).

* Naval Research Laboratory Contribution Number NRL/JA/7320-04-4.

Corresponding author address: W. Erick Rogers, NRL Code 7322, Bldg. 1009, Stennis Space Center, MS 39529.
E-mail: rogers@nrlssc.navy.mil

Both are known as "third-generation" wave models. Recent reviews of the Navy's operational global wave models can be found in Jensen et al. (2002) and Wittmann (2001). It is expected that future development and updating of the Navy's WAM code will be much less active than that of the WW3 code. Thus, in this paper, hindcasts are performed only with the WW3 model.

a. *Prior wave model evaluations*

The FNMOC WAM4 model (since replaced by WW3) was compared to models at other operational centers by Bidlot et al. (2002). Two earlier references on Navy global wave modeling are Clancy et al. (1986), Wittmann and Clancy (1993), and Wittmann et al. (1995). The WW3 global implementation at the National Centers for Environmental Prediction (NCEP) is evaluated in Tolman et al. (2002).

In prior investigations of both operational U.S. Navy global wave models [WAM and WW3; see Rogers (2002) and Rogers and Wittmann (2002)], it was determined that the dominant error in those models (during the periods of January 2001 and January–February 2002) was very likely caused by the inaccuracy of the forcing fields from the operational global atmospheric model NOGAPS, in particular a negative bias in predictions of high wind speed ($U_{10} > 15$ m s^{-1}) events by that model. Bias associated with the wave model itself (internal error) was believed to be only secondary.

b. *The operational meteorological product*

For wind forcing, both of the Navy's global wave models use wind vectors from the Navy Operational Global Atmospheric Prediction System [NOGAPS; see, e.g., Hogan and Rosmond (1991) and Rosmond et al. (2002)]. We will not attempt to describe the many features of NOGAPS here, but will limit discussion to relevant features of the model. Prior to August 2002, NOGAPS used the Emanuel cumulus parameterization—described in Emanuel and Zivkovic-Rothman (1999) and Teixeira and Hogan (2002)—and was run at T169L24 resolution (~80 km horizontal resolution at midlatitudes, 24 vertical levels).

The operational NOGAPS model was significantly modified during 2002. Teixeira and Hogan (2001) state that the Emanuel cumulus scheme in NOGAPS likely produces a negative bias in surface winds and suggest an improvement that was implemented in the operational NOGAPS in August 2002. According to Teixeira and Hogan (2001), the new cloud scheme reduces the surface wind bias. The horizontal and vertical resolution of NOGAPS was upgraded in September 2002, from T169L24 to T239L30 (~50 km horizontal resolution at midlatitudes, 30 vertical levels), which may further reduce negative bias in the surface winds.

c. *Outstanding questions*

Major improvements in operational surface wind forcing fields usually lead to significant (and sometimes dramatic) improvements in the operational wave model results. In this paper, we demonstrate one such case. This result is perhaps obvious (or at least, anticipated) enough that a demonstration of such might seem banal. The more interesting questions one might ask are the following:

- If two competing forcing fields are less dissimilar in skill (say comparing products from two operational centers, or comparing analysis fields versus forecast fields), does the more accurate field necessarily yield better wave model results?
- How might the metric for accuracy be different for a wave modeler than, for instance, a circulation modeler? For example, how important is random error relative to bias error?
- If we can identify a scenario where a wave model's representation of physics (generation, dissipation, and nonlinear interactions) is likely to be the primary source of error, is the wave model bias positive or negative? How does the answer depend on the frequency–wavenumber range considered, or perhaps the geographic location?

The purpose of this paper is to answer these questions. This will be done using hindcasts that are designed specifically for this purpose.

d. *Outline*

The remainder of this paper is structured as follows: section 2 is a description of the operational (FNMOC) wave model. In section 3, a review of types of model errors is given. This review provides the basis for tests that are used in the evaluation of the FNMOC global wave predictions. These tests are presented in section 4, along with additional description of the method used in this study to answer the questions above, and description of the hindcast design. Section 5 describes the hindcast results. Discussion is given in section 6, and conclusions (corresponding to the three questions above) are summarized in section 7.

2. Model description

WAVEWATCH-III is phase averaged. This implies that output from the model is relevant on time scales longer than the waves themselves, and that computational geographic resolution can be much greater than one wavelength. The governing equation of WW3 is the

action balance equation, which in spherical coordinates is (Tolman 2002a)

$$\frac{\partial}{\partial t}N + (\cos\phi)^{-1}\frac{\partial}{\partial\phi}\dot{\phi}\cos\phi\, N + \frac{\partial}{\partial\lambda}\dot{\lambda}N + \frac{\partial}{\partial k}\dot{k}N + \frac{\partial}{\partial\theta}\dot{\theta}_g N = \frac{S}{\sigma}, \tag{1}$$

where t is time; λ is longitude; ϕ is latitude; θ is wave direction; N is the wave action density spectrum, described in five dimensions (λ, ϕ, k, θ, t); k is the wavenumber; the overdot symbol denotes the wave action propagation speed in (λ, ϕ, k, θ) space; σ is relative frequency;[1] and S is the total of source/sink terms (these are often referred to as the "physics" of a wave model). Wave action density is equal to energy density E divided by relative frequency ($N = E/\sigma$). If currents are not considered, which is presently the case at NAVO, FNMOC, and NCEP, then the action density and energy density conservation equations are essentially identical.

In deep water, S is dominated by three terms: $S \approx S_{in} + S_{nl} + S_{ds}$, input by wind (which can be negative in the case of WW3), four-wave nonlinear interactions, and dissipation, respectively. The physics of WW3 are described in Tolman and Chalikov (1996), with minor refinement of the Tolman and Chalikov physics being described in Tolman (2002a). For the most part, the physical formulations of this model are based on earlier works, some of which are not referenced herein.

The life cycle of a wave train can be divided into a "growth" or "generation" stage and a propagation stage. During the growth stage, all three source/sink terms are important. To accurately predict wave growth, all three terms must be skillful, or at least must be tuned such that shortcomings in any one term will tend to be compensated by other term(s). At the propagation stage, once swells are sufficiently dispersed such that the wave steepness is small, nonlinear interactions are insignificant. Also, the ratio of wind speed to wave phase velocity is rarely high enough to transfer momentum to longer swells (i.e., most often, $S_{in} \leq 0$). Thus, in that case, only attenuation is important. In WW3, attenuation is represented by combined S_{in} and S_{ds} (both negative).

WW3 uses finite-differencing methods to approximate the partial differential equation given in (1). The implementation at FNMOC uses the higher-order (more accurate) approximations available in WW3. For more detail regarding WW3, see Tolman (2002a) and references therein.

[1] If currents are present, this is the frequency measured in a frame moving with the current.

The FNMOC WW3 model was run at 1° resolution prior to October 2002. After October 2002, it has been running at 0.5° geographic resolution. Wave spectra in the global model are at a resolution typical of operational third-generation models.[2]

3. Model errors: A review

a. Numerics and resolution

Underprediction of swell energy has long been a problem in the Navy's global wave models. When WAM4 was the only global wave model at the Navy (prior to 2001), the underprediction was in informal communications often ascribed to the relatively primitive numerical techniques used in WAM4. Bender (1996) conducted a validation study of WAM cycles 2 and 4 in the Southern Hemisphere (performed for the Australian Bureau of Meteorology) and concluded that "the first-order upwinding propagation numerics of WAM is clearly responsible for excessive dissipation of wave energy—in particular, swell." This reinforced the belief that numerical inaccuracy of the first-order propagation scheme of WAM4 was the root cause of underpredicted swells in the Navy global WAM4. However, Wittmann and O'Reilly (1998) and Rogers (2002)—through the use of a great circle wave ray-tracing tool developed by Dr. W. C. O'Reilly (Scripps Institution of Oceanography)—demonstrated that the diffusion associated with the first-order scheme of WAM is unlikely to be a primary source of negative bias in the Navy's global WAM4 implementation, even if only older swells (which are the wave frequency ranges most affected by diffusion) are considered. In fact, this is consistent with the nature of the numerical schemes used by the models: they are mass conserving, so the schemes do not dissipate energy and cannot be directly responsible for negative bias [though they might be indirectly responsible, e.g., in conjunction with blocking by landmasses, which can lead to local bias; e.g., Rogers et al. (2002)]. At a given geographic location, some spectral components may have significant errors associated with propagation numerics, while another spectral component may have much smaller error, or error of opposite sign (between components). Thus, at that location, the effect of numerical geographic propagation error on wave height (i.e., the integrated wave spectrum) will tend to be smaller than its effect on individual spectral components. The WW3

[2] Twenty-four directional bands (thus, 15° directional resolution) and 25 frequency bands (with logarithmic spacing, the interval being 10% of the frequency, and the first frequency being 0.0418 Hz) are used.

model provides the option of employing higher-order propagation numerics (and other, related improvements), so the impact of this issue is diminished even further in the case of that model. The ray-tracing method also eliminates problems with geographic and spectral resolution that manifest during swell propagation modeling. Thus these studies can also be taken as evidence that geographic and spectral resolutions were not a primary source of bias. However, it should be pointed out that geographic resolution *is* expected to play an important role in some locations (see Tolman 2003). The conclusions about bias may also be true with regard to rms error: though one might expect rms error to be more sensitive to numerics and resolution, we have yet to see in our extensive studies any case in which improved propagation methods yield significant reduction in rms error *except* in cases of special local effects (e.g., near islands). Supporting information and discussion of other numerics/resolution issues can be found in Rogers (2002). [The term numerical diffusion (or just diffusion) is used in this paper to describe the unintended spreading or smearing of wave energy due to discretization of a continuous problem, more specifically due to even-ordered truncation error terms in the governing equation finite-differencing associated with propagation.]

WW3 uses specialized methods to deal with the problems associated with coarse spectral discretization, namely, that of Booij and Holthuijsen (1987). Additional refinement and improvement to propagation are new features of WW3, version 2.22 (see Tolman 2002b; Tolman 2003). At the time of this writing, this version is operational at FNMOC, but not all of the new features have been activated (implementation is forthcoming).

b. *Source/sink terms*

Inaccuracies that occur during swell dispersion may be significant in WW3. ["Dispersion" is used herein to describe the process of the dispersion of waves of different velocity and direction of propagation.[3]] Given accurate forcing, the total energy of windseas is very often well predicted by the model physics, as one might expect from a well-behaved model with low-order tuning. However, the frequency and in particular the directional distribution of this energy have not been extensively validated (the same might be said about the WAM model). Thus, predictions of details of wave spectra are typically less accurate than predictions of total energy (wave height). Inaccuracy in the spectral distribution of low-frequency energy leads to inaccuracies as the low-frequency windsea disperses as swell. Given long propagation distances, this error can become large relative to the height of the swells. This does not always have a profound impact on rms error, as older swells often constitute a small portion of the wave spectrum at any given time/location; in these cases, the effect on wave height (total energy) predictions will be small (in other words, the error is usually masked by local windsea). But in climates dominated by older swells (e.g., the Tropics), the effect may be relatively significant.

The physics of swell attenuation is not expected to be accurate in third-generation models. Again, this is usually evidenced by poor agreement with altimeter wave heights in the Tropics. However, it is difficult to isolate the effect of swell attenuation from other problems that produce similar underprediction (like, say, inaccuracy in frequency–directional distribution). One might identify specific cases where a swell field passes two buoys at different stages in the swell field's life cycle, but that too can be troublesome since one buoy might measure the geographic center of a swell field, while another might measure the outer edge; considerable care is required.

c. *Wind forcing*

Inaccuracies in the wind forcing used by a global wave model are another source of error. There have been several studies dealing with the accuracy of atmospheric predictions from the perspective of the wave modeler. In fact, it is a standard practice to evaluate wind field accuracy alongside wave field accuracy. Such studies include Komen et al. (1994), Cardone et al. (1995), Cardone et al. (1996), Khandekar and Lalbeharry (1996), Janssen et al. (1997), and Tolman (1999). The Cardone et al. (1996) study deals with several wave models, including WAM4. Their analysis is in terms of total wave height (rather than particular frequency bands), but they focus on extreme events that are characteristically dominated by low-frequency energy. They find that, given accurate forcing, wave model bias is very small for wave heights under 12 m, and that in operational nowcast/forecast systems, wind forcing is the dominant source of error.

Rogers (2002) compares surface winds from NOGAPS and NCEP analyses to National Aeronautics and Space Administration (NASA) Quick Scatterometer measurements (QuikSCAT; PODAAC 2001) in the northeast Pacific during January 2001 and in the South Pacific during July 2001. It is found that both

[3] Of course, "numerical dispersion" is an appropriate term for the odd-ordered truncation error terms in the governing equation finite differencing associated with propagation. We do not discuss this type of numerical error specifically herein, but include it in the more general "propagation error."

analyses tend to be biased low at high wind speeds, but the bias is relatively slight with the NCEP analyses and quite significant in the NOGAPS analyses, particularly in the northeast Pacific comparison. Rogers and Wittmann (2002) make similar direct wind comparisons, but over the globe for 1 January–8 February 2002; these similarly suggest that strong surface wind events in the NOGAPS analyses were biased low. The negative bias in high wind speeds of NOGAPS observed by Rogers (2002) and Rogers and Wittmann (2002) during January 2001 and January–February 2002 were presumably due to the Emanuel cumulus scheme used at that time (mentioned above).

During those studies, it became apparent that the extensive coverage of the QuikSCAT dataset (90% or more of the ocean surface every day)—together with its directional capability—make it possible to derive "snapshot" wind fields using that dataset, which could be used to force a wave model. Thus, Rogers (2002) and Rogers and Wittmann (2002) also made indirect comparisons of surface wind fields: competing hindcasts forced by (a) NOGAPS analyses, (b) these NOGAPS fields supplemented with QuikSCAT measurements, or (c) NCEP global wind analyses. Rogers (2002) investigates the sensitivity of the January 2002 hindcast case to wind forcing using time series comparisons of low-frequency wave energy. In the January 2002 hindcasts, Rogers (2002) looked at the wave climate in the North Pacific, so it was essentially a hindcast of the low-frequency energy generated by the strong extratropical storms typical of this time and location. Rogers and Wittmann (2002) take an alternate tack by including not only local comparisons but also regional and global comparisons. They use the TOPEX/Poseidon altimeter data [for description see Fu et al. 1994 and the *European Remote Sensing Satellite-2* (*ERS-2*)]. The comparison thus differs further from Rogers (2002) insofar as it is of wave height (or total wave energy), which is the quantity that is inferred from altimeter measurements. The results of the indirect comparisons were consistent with the direct comparisons. The NOGAPS-forced model and NCEP-forced models are both biased low relative to buoy data, whereas the models forced by wind fields supplemented with scatterometer data performed very well.[4] Moreover, Rogers (2002) concluded that wind forcing was likely to be the dominant source of error in operational low-frequency energy predictions and that given accurate forcing, both WAM and WW3 predict young low-frequency energy rather well, consistent with observations of Cardone et al. (1996).

[4] In these studies, the blended NOGAPS–QuikSCAT fields were not validated against independent data; the systematic error in the QuikSCAT measurements were assumed small relative to those in NOGAPS. This leads to increased uncertainty in the conclusions. In the present study, all wind fields are validated (section 5a).

d. Other external errors

With a wave model it is possible to have other sources of external errors (besides wind forcing). We do not expect that any of these sources of error could be significant in a global wave model; we include this discussion for the sake of completeness.

One external error is imperfect knowledge of bathymetry, coastline, ice edge, and so forth. This type of error is expected to be very small at the global scale, since the real challenge is not to know the bathymetry, but rather to resolve the known bathymetry with the computational grid.

Another external error is boundary forcing. This is extremely important to wave models in general, but is not relevant to a global model for obvious reasons.

A third and fourth type of external errors are from poor specification or nonspecification of currents and/or air−sea temperature differences (stability). These are also not expected to have a significant impact in global applications.

4. Method

As we have discussed, total error in global wave model predictions comes from two sources: external (wind forcing) and internal (wave model physics, numerics, resolution). Our purpose here is to evaluate internal errors. The major challenge is that during wave model validation, one can discover a great deal about the total error (from comparison with wave observations), but not the apportionment of external and internal errors. Further, it is useful (though not particularly easy) to quantify bias associated with wave model source/sink terms. It is possible to construct tests, with the intent of addressing this challenge. In section 4a, we present three tests, two of which are applied in this study. In section 4b we describe the global wave model hindcasts, to which these two tests will be applied. In section 4c, we describe the wind fields used to force these hindcasts, with a summary of evaluation of bias in these wind fields. In section 4d, we describe the wave observations used as ground truth in this study and the metrics used.

a. Evaluation method (conditional interpretation)

There are three condition–interpretation pairs constructed to learn more about the errors in wave model predictions.

1) If a model is forced with a wind field that contains *zero* bias, then any bias observed in energy predictions from a wave model forced by these wind vectors implies bias associated with the wave model itself.
2) If a model is forced with a wind field with a bias of known sign, and bias of opposite sign is observed in energy predictions from a wave model forced by these wind vectors, this implies bias associated with the wave model itself.
3) If hindcasts and wave model–data comparisons are chosen such that the bias from numerics and resolution is *nonexistent*, then the bias in the wave model itself (i.e., bias not associated with wind field accuracy) is associated with the model source/sink term parameterizations.

Of course, our knowledge is not absolute, so we must recast these tests in an approximate form (with "a" to indicate "approximate"):

1a) If a model is forced with a wind field that contain *small* bias, then *nontrivial* bias observed in energy predictions from a wave model forced by these wind vectors implies a *probable* bias associated with the wave model itself.
2a) If a model is forced with a wind field with a bias of known sign, and *nontrivial* bias of opposite sign is observed in energy predictions from a wave model forced by this wind field, this implies a *probable* bias associated with the wave model itself. (No conclusions are drawn if the bias is of the same sign.)
3a) If hindcasts and wave model–data comparisons are chosen such that the bias from numerics and resolution is *small*, then the *nontrivial* bias in the wave model itself (i.e., internal bias) *is probably associated* with the model source/sink term parameterizations.

The approximate tests are subjective, depending on the evaluator's idea of the terms small, trivial, and reasonably well known. The term probably is also imprecise. Test 2a is somewhat more objective than test 1a, as it does not require definition (or proof) of small bias in a wind field, and does not require careful, separate evaluation of the sensitivity of the wave model to wind field bias. Thus, in this study, we apply test 2a but not test 1a. Test 2a is quite straightforward. Note, however, that in cases where the wind field bias is very large, test 2a will not be useful, since the bias from external error overwhelms any bias from the internal errors. Test 2a makes it apparent that it is useful to have two alternate wind forcing fields, with bias of opposite sign; this is one of the primary motivations for including alternate wind analyses in our hindcasts (see section 4c).

Test 3a requires significant further explanation. How does one ensure that errors from numerics and resolution are small? In our case, we make the following arguments: Recall from the review (section 3) that previous studies regarding the effects of propagation error found that these factors do not have a significant effect on wave model bias in unsheltered areas, even if very old (e.g., greater than 8 days old) swells are considered, and even when the first-order propagation scheme of WAM4 is employed. (This is the expected result, given the nature of propagation errors.) Hindcasts in the present study are limited to months corresponding to winter in the Northern Hemisphere, and wave observations employed are strictly in the Northern Hemisphere (U.S. Atlantic coast and U.S. Pacific coast). We know from climatology[5] that observed wave energy at these times/locations is dominated by windsea and young swells (0–5 days old). Error associated with numerics and resolution will tend to accumulate as swell propagates. Seas and young swells would have propagated a shorter distance and have therefore accumulated less of such error than older swells used in the reviewed literature. Further, we apply the WAVEWATCH-III model, which employs a higher-order propagation scheme, thus further reducing the impact of numerics and resolution.

In this study, the conclusions from the application of test 3a are made more specific by choosing locations where finite water depth physics (e.g., bottom friction) can be assumed small (see section 4d).

b. Description of wave model hindcasts

These hindcasts differ from those of Rogers (2002) in that they are of more recent time periods (thus reflecting recent changes to the NOGAPS model), of longer duration, and are limited to the wintertime in the Northern Hemisphere. Two hindcast time periods are used: winter 2001/02 (0000 UTC 1 December 2001–2100 UTC 3 March 2002) and an identical time frame for the winter of 2002/03. For each winter, two wave model hindcasts are forced by two different wind fields: NOGAPS and blended NOGAPS–QuikSCAT data. Thus, four hindcasts are performed. Winds are prescribed on a 3-h interval, except for during December 2001, for which a 6-h interval is used [corresponding to

[5] We have verified that this statement about the climate is accurate. To do this, we used 1) directional wave spectra inferred from NDBC buoy 46042 observations and 2) animated time series of wave height fields (for the North Pacific) from one of the model hindcasts. We do not dispute that swell from the Southern Hemisphere occurs at these times/locations, but the evidence suggests that their impact on our time series is trivial.

the interval that the fields were available in the Naval Research Laboratory (NRL) archives].

All four hindcasts were set up similarly to the global implementation of WW3 at FNMOC prior to October 2002 (see section 2). One degree of geographic resolution is used for all hindcasts, as the impact of geographic resolution is not studied herein (to do so, one would need to run two hindcasts that are identical except for their geographic resolution).

c. *The wind fields*

1) Description of the wind fields

Two of the hindcast wind fields are taken from NOGAPS analyses. For the other two fields, we supplement the NOGAPS analyses with wind vectors inferred from scatterometry. Specifically, we use QuikSCAT level 2B (henceforth denoted L2B) data provided by the Jet Propulsion Laboratory's Physical Oceanography Distributed Active Archive Center (PODAAC). See PODAAC (2001) for a description of this dataset.

To summarize, we have four simulations, with one major difference between them (the wind field applied) and one minor difference between them (year-to-year variation in climate, which might cause minor discrepancies between error metrics calculated for a winter 2001/02 hindcast and that for a winter 2002/03 hindcast). The four wind fields are:

1) NOGAPS analyses from a time period prior to the August 2002 modifications to NOGAPS (winter 2001/02);
2) wind fields of 1) above, supplemented with QuikSCAT data;
3) NOGAPS analyses from a time period after the August 2002 modifications to NOGAPS (winter 2002/03); and
4) wind fields of 3) above, supplemented with QuikSCAT data.

This leads us to an important, obvious point, which is that since the only major difference in these simulations is the wind field, the comparison of one hindcast to another reflects sensitivity of the wave model to the wind field.

We create the blended NOGAPS–QuikSCAT wind fields using a relatively simple method: at a particular time (corresponding to the time of the snapshot map, which is calculated at 3-h intervals), longitude, and latitude (corresponding to a point in the forcing grid), the wind vector is calculated using the following logic: *If a QuikSCAT measurement is nearby (in time and space), we use that measurement. Otherwise, we use the NOGAPS value for that time/location.*

"Nearby" geographically is simply defined as falling within a particular model grid cell. The definition of whether a measurement is nearby in temporal space is more subjective. If this temporal window is very large (e.g., ± 12 h), this is expected to increase random errors in the wave field. Three different windows were tested and the "± 6 h (12-h window)" criterion for data usage was chosen, since that provides reasonably good coverage of the ocean's surface. Figure 1 shows an example of the global coverage obtained with this 12-h window. No smoothing procedures (see, e.g., Chin et al. 1998) were performed, since a wind field with large degree of nonuniformity and nonstationarity does not present problems (e.g., with stability or consistency) for a large-scale phase-averaged wave model. For similar reasons adjustment for meteorological consistency (e.g., via adjoint techniques with a dynamical atmospheric model) was not necessary. For the QuikSCAT data, we discarded all data flagged for possible quality problems (such as rain presence), with one exception: the flags relating to very low (less than 3 m s^{-1}) and very high wind speeds (greater than 30 m s^{-1}) were not considered. The filtering was thus identical to that used in the validation of L2B data by Ebuchi et al. (2002).[6]

2) Validation of the wind fields

Test 2a (and also test 1a, which we do not apply) in the previous section requires that bias in the wind fields are well understood. In this section, we therefore validate the wind fields used to force the global wave model hindcasts.

The wind fields used to force the hindcasts are compared to in situ wind data.[7] Due to constraints on manuscript length, this validation cannot be included here, but is presented in a separate publication (Rogers et al. 2004). For quantitative results, we refer the reader to that report; here we summarize the results qualitatively.

The "bias" in the wind fields was evaluated by Rogers et al. (2004) for individual wind speed bins. The bias in each bin is weighted according to its expected effect on the wave model: for the sake of simplicity, the Pierson–Moskowitz (PM) wave height (Pierson and

[6] The decision to omit flagged data was made based on the results of the wind forcing validation (Rogers et al. 2004). In earlier hindcasts, we did not discard quality-controlled flagged data. This was motivated by a desire to preserve measurements in the vicinity of storms (where the accuracy of the meteorological models is most critical to the wave model). Thus, greater coverage was favored over greater quality. However, in our validation of the wind forcing fields, we find that this tactic is not justified.

[7] Buoy data are used. To use QuikSCAT data to evaluate the NOGAPS–QuikSCAT fields would have been problematic for obvious reasons.

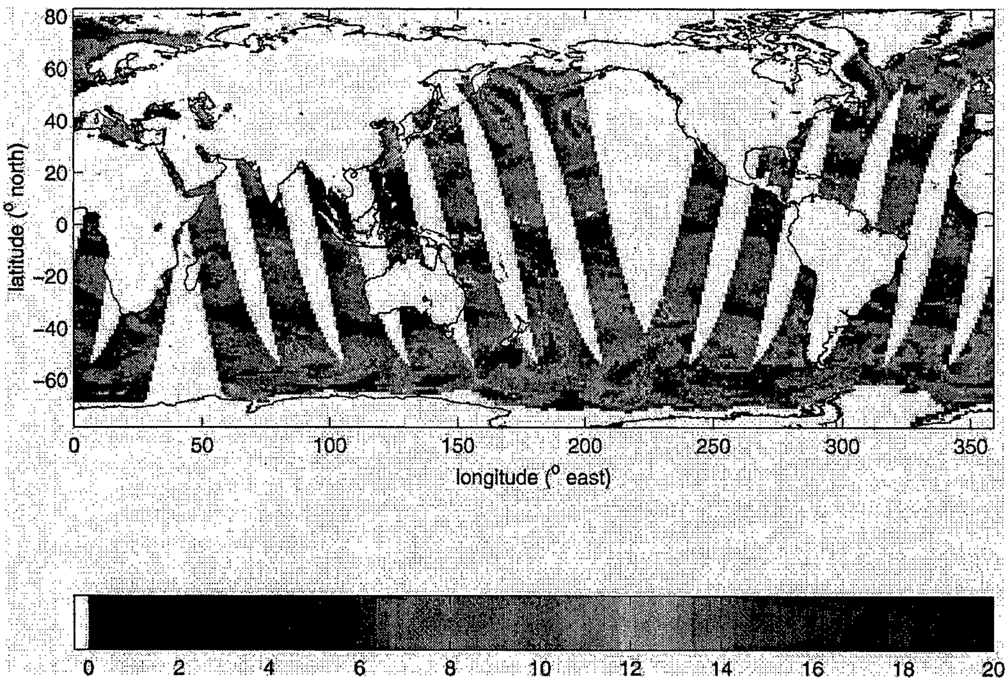

FIG. 1. Wind field map created from QuikSCAT data (wind speed, m s^{-1}). The time of the "snapshot" shown is 2100 UTC 3 Mar 2003. In practice, NOGAPS forcing would be used to fill in gaps (where data are not used), but are shown as blank areas here. The PODAAC L2B QuikSCAT data within 6 h of the snapshot time (before or after) are included in the snapshot.

Moskowitz 1964) is used. This can be interpreted as a weighting according to the square of the wind speed.

This leads to the following conclusions:

1) Of the four forcing fields, the 2001/02 NOGAPS has the most severe bias.
2) For both time periods, the rms error of the blended NOGAPS–QuikSCAT fields is higher than that of the NOGAPS fields.
3) The 2002/03 NOGAPS field has the smallest bias. The negative bias at high wind speeds is greatly reduced compared to the previous winter. [Note that this improvement corresponds to upgrades to the resolution and cloud parameterization of the NOGAPS model (see, e.g., Teixeira and Hogan 2001).]
4) The apparent bias in the NOGAPS fields (for both time periods) is negative and thus is expected to result in an underprediction of wave energy by a "perfect" wave model. The negative bias is primarily at 10-m wind speeds greater than 12 m s^{-1}.
5) The apparent bias in the blended NOGAPS–QuikSCAT fields is positive and thus is expected to result in an overprediction of wave energy by a "perfect" wave model. The positive bias is primarily at 10-m wind speeds greater than 15 m s^{-1}.
6) The apparent bias in the blended NOGAPS–QuikSCAT fields is remarkably similar for the two time periods. This might be taken as an indication of the robustness of the method.

d. *Wave model ground truth and metrics used*

In the discussions to follow, the wave hindcasts forced with the NOGAPS wind fields are denoted "NF model" and the wave hindcasts forced with the blended NOGAPS–QuikSCAT fields are denoted "QNF model."

National Data Buoy Center (NDBC) buoy spectra are used as ground truth for model evaluation. In comparison to these measurements, the first 6 days of the simulations were omitted to accommodate for model spinup. Wave spectra from the simulations were saved for a number of locations, of which seven are used for comparisons to data.

The locations of the observations are chosen such that the dominant wave climate consists of windsea and younger swells (being defined here as swells 1–5 days old). Thus, the impact of (a) numerical error (e.g., diffusion) and (b) spectral resolution (e.g., via the garden sprinkler effect; Booij and Holthuijsen 1987) are both minimized, allowing more definitive conclusions on the causes of error (see discussion in section 4a). Similarly,

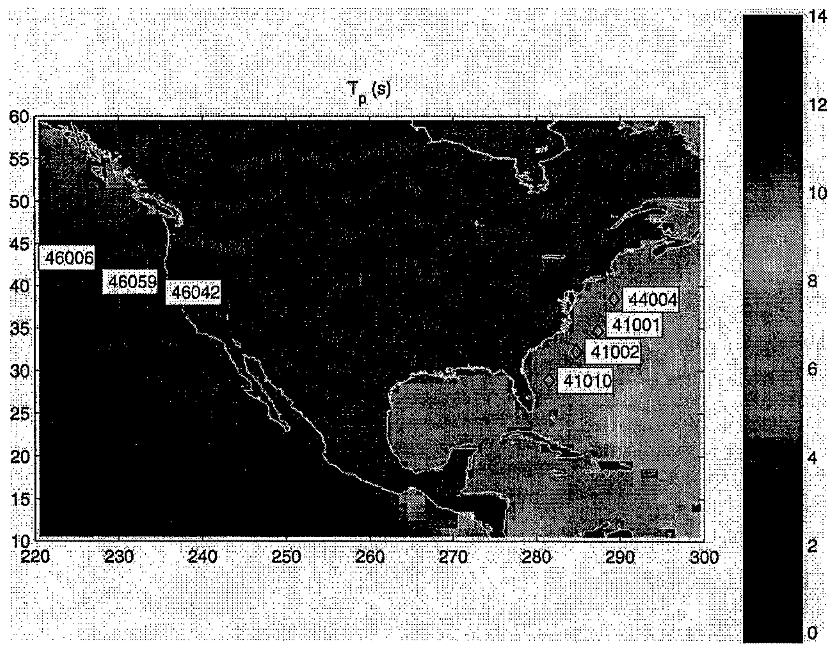

FIG. 2. Locations of NDBC buoys used in comparisons to model output are shown. Color shading indicates representative distribution of peak wave period (s).

the buoy locations are chosen with a preference for deeper water locations to minimize the effect of wave–bottom interactions. Preference was also given to predominantly unsheltered locations to avoid the effects of geographic resolution inasmuch is possible. Thus—applying test 3a above—the internal wave model bias is attributable primarily to the deepwater physics of the model (wind input parameterization, whitecapping and swell attenuation parameterization, and nonlinear interactions).

The seven buoy locations are indicated in Fig. 2. A representative peak wave period distribution is also shown in Fig. 2, with the intent of showing some aspects of the variation of the wave climate at the seven locations.[8] Waves at the three Pacific Ocean buoy locations tend to be generated over greater fetches than is the energy at the four Atlantic Ocean buoy locations. This is partly attributable to the general trend of extratropical weather systems traveling from west to east, creating dynamic fetch situations more often in the northeast Pacific versus the northwest Atlantic. Thus, the northeast Pacific Ocean can be characterized as swell dominated (and during the winter time, young-swell dominated), and the East Coast can be characterized as a more even mixture of seas and young swells.

Model output is provided at 1-h intervals and is collocated with the buoy data via bilinear interpolation. NDBC provides buoy data at hourly intervals. All buoy data used in calculations of bias and rms error have been subjected to a 3-h running average. This low-pass filtering is to reduce nonstationarity in low-frequency bands.[9] The 3-h interval is chosen—as opposed to a longer interval, which is expected to give results more favorable to a wave model, which tends to be smooth—after consideration of the interval of the wind forcing fields (in these hindcasts and operationally).

In this evaluation, there is a departure from the traditional metrics based on wave height and peak period. Wave height (or total energy) is used, but peak period is not. Instead of peak period,[10] we look at statistics for

[8] The representative peak period distribution here is calculated as the time average of all distributions of peak wave period for the time period 7 December 2002–3 March 2003 from the QNF model.

[9] This nonstationarity is partly attributable to the relatively short 20-min data interval typically used by NDBC and difficulty of measuring low-amplitude swells via accelerometer.

[10] Peak period is often used because it is easy to understand and has an objective definition. Yet it is a fairly useless metric in cases of multiple peaks of similar magnitude. There is also a problem of minor details of model spectral shape having a large impact on the peak period (H. Tolman 2004, personal communication). Mean period is preferable, though it suffers from highly subjective definition. Inspection of frequency bands is a bit more work, but can provide knowledge of behavior not apparent from these bulk parameters.

four frequency bands. (Model frequency bins do not fall neatly within the bands that we use, so simple linear interpolation of the one-dimensional spectra is performed.) It is useful to look at the model in such a way, since biases at different frequency bands often tend to cancel each other when the spectrum is integrated to calculate total wave energy. Note that the 0.04–0.40-Hz range comprises essentially the entire wind–wave spectrum; the wave height calculated from integrating the wave spectrum 0.04–0.40 Hz can be considered equivalent to the significant wave height. The other wave heights (integrated over different frequency ranges) can be considered "partial" wave heights (this quantity was chosen rather than variance, since wave height has a more visceral quality). The separation into four frequency bands is *not* intended as a method of sea–swell separation, though one can make educated judgments about the constituency of certain frequency bands: for example, it is improbable that much 4–6-day-old swell energy exists in frequencies greater than, say 0.10 Hz.

It is useful to distinguish between systematic bias and random error. Random error, typically less cause for alarm to a modeler than systematic error, is also much more difficult to trace to a consistent source. In the discussion to follow, we define error as being "predominately random" if the rms error is greater than three times the magnitude of the bias. (Random error is more precisely defined using the "standard deviation of the error," which we do not include here. However, when bias is small, rms error approximates the standard deviation of the error.) The correlation coefficient is included in tables. This is calculated as

$$r = \frac{\sum_{i}^{n} (H_{\mathrm{obs},i} - \overline{H}_{\mathrm{obs}})(H_{\mathrm{hc},i} - \overline{H}_{\mathrm{hc}})}{\sqrt{\sum_{i}^{n} (H_{\mathrm{obs},i} - \overline{H}_{\mathrm{obs}})^2 \sum_{i}^{n} (H_{\mathrm{hc},i} - \overline{H}_{\mathrm{hc}})^2}},$$

where r is the correlation coefficient, H is the wave height, subscripts obs and hc denote observed and hindcast, and n is the number of collocated points. The utility of the correlation coefficient is to distinguish between cases where bias and rms error are low due to good model performance versus cases where bias and rms error are low due to generally low wave heights in the time series.

5. Results

Tests 2a and 3a introduced in section 4a can be restated in a form specific to the hindcasts herein. From test 2a, we see the following:

- *in the NF model results, positive bias is predominately due to the wave model (i.e., internal bias) and*
- *in the QNF model results, negative bias is predominately due to the wave model (i.e., internal bias).*

As discussed above, we expect that bias associated with propagation error (e.g., numerics, resolution) does not have a significant impact on bias, and that finite-depth source sink terms (i.e., wave–bottom interaction) contribute little to bias at the comparison locations. Thus, from test 3a, we expect that *internal bias in these hindcasts is (to first order) bias associated with the wave model's deepwater source/sink term parameterizations.* These three tests (shown in italics) are applied in sections 5a and 5b.

Model results are tabulated in Tables 1 and 2 and example time series are shown in Figs. 3a and 3b. There is a striking difference between these two figures: note the dissimilarity between the two forcing methods in the 2001/02 hindcasts and the similarity of the two forcing methods in the 2002/03 hindcast. This clearly shows that—viewed through the filtering effect of the wave model—the NOGAPS fields have become much closer to the scatterometer measurements in the intervening period. There is considerable remaining error however, mostly underpredictions by the models.[11]

Due to the large number of hindcast–location–frequency band combinations, further discussion will be limited to the error statistics shown in the tables. [However, readers seeking more time series comparisons (each row in the tables corresponds to a separate plot) can find them as supplementary material online at http://dx.doi.org/10.1175/waf882.s1.]

Originally, error metrics were listed by individual buoys, but during the evaluation, it became apparent that trends are fairly consistent for different buoys of the same ocean. Thus, in Tables 1 and 2, error metrics are more concisely presented using averages calculated by buoy groups (northwest Atlantic and northeast Pacific). The full listings are given in the supplementary material.

a. Winter of 2001/02

During the winter of 2001/02, rms error in the NF model is consistently higher than is the rms error with the QNF model. In all basin-averaged bias calculations, the magnitude is smaller with the QNF model. In most cases, it is dramatically reduced. Thus, from the winter

[11] The underprediction by the QNF model in early December 2001 is particularly noticeable. This is probably due to a set of nearby and relatively small-scale storms not well measured by the QuikSCAT instrument.

TABLE 1. Results for winter of 2001/02. Error measures of WW3 hindcast wave heights are shown, with NDBC buoy data used as ground truth. "Partial wave height" is calculated from the variance (i.e., energy) of the wave spectrum over a frequency range defined by lower and upper bounds f_1 and f_2: $H_{m0,\text{partial}} = 4\sqrt{v_{\text{partial}}}$, where $v_{\text{partial}} = \int_{f_1}^{f_2} E(f)\,df$ is the "partial variance." Also, E is spectral density, f is frequency. Bias refers to the mean error. Rmse is root-mean-square error. The rows corresponding to total wave height H_{m0} are highlighted in boldface.

	f_1 (Hz)	f_2 (Hz)	Bias (m) NOGAPS forcing	Bias (m) NOGAPS–QSCAT forcing	Rms error (m) NOGAPS forcing	Rms error (m) NOGAPS–QSCAT forcing	Correlation coef NOGAPS forcing	Correlation coef NOGAPS–QSCAT forcing
Atlantic	**0.04**	**0.40**	**−0.48**	**−0.18**	**0.69**	**0.50**	**0.92**	**0.91**
	0.04	0.06	−0.03	0.00	0.10	0.09	0.43	0.55
	0.06	0.08	−0.17	−0.07	0.40	0.33	0.53	0.66
	0.08	0.12	−0.55	−0.27	0.86	0.61	0.79	0.84
	0.12	0.40	−0.16	0.01	0.38	0.35	0.88	0.89
Pacific	**0.04**	**0.40**	**−0.49**	**−0.06**	**0.81**	**0.52**	**0.90**	**0.94**
	0.04	0.06	−0.56	−0.28	0.80	0.51	0.76	0.83
	0.06	0.08	−0.47	−0.02	0.76	0.45	0.79	0.89
	0.08	0.12	−0.11	0.04	0.50	0.41	0.87	0.91
	0.12	0.40	−0.09	0.01	0.32	0.27	0.90	0.93

2001/02 hindcasts, one can conclude that there is a clear advantage to supplementing wave model forcing fields with QuikSCAT data in this manner. This result is especially remarkable if one considers that in the wind field validation (Rogers et al. 2004, summarized above), it was found that the rms error of the NOGAPS fields is *less than that of the QuikSCAT–NOGAPS fields*. This suggests that bias of winds at moderate and high winds speeds, as a metric, is much more relevant to wave predictions than is wind speed rms error.

It is useful to observe the residual error after the bias in the wind forcing is reduced (i.e., error in the 2001/02 QNF model results). With regard to the entire 0.04–0.4-Hz range, the bias is negative. This overall bias is of moderate magnitude at the northwest Atlantic buoy locations. The negative bias at these locations is mostly limited to the dominant frequency bands, 0.06–0.12 Hz. Most of the remaining error appears to be random in nature. At the northeast Pacific locations, the overall bias is small, but the negative bias at the lowest frequencies (0.04–0.06 Hz) is quite large. Again, most of the remaining error appears to be random in nature.

Using test 2a described in section 4a, we can say that some bias is probably attributable to the wave model itself, and from test 3a, it is (more specifically) likely associated with deepwater source/sink term. Specifically, the following biases are seen:

- in the Atlantic, a negative bias in total wave energy;
- in the Atlantic, a negative bias at 0.06–0.12 Hz; and
- in the Pacific, a negative bias at 0.04–0.06 Hz.

Since the negative bias in the NOGAPS wind forcing is so strong, we cannot detect positive bias attributable to the wave model for the winter 2001/02 hindcast using

TABLE 2. Results for winter of 2002/03. Error measures of WW3 hindcast wave heights are shown, with NDBC buoy data used as ground truth. See Table 1 for definition of terms.

	f_1 (Hz)	f_2 (Hz)	Bias (m) NOGAPS forcing	Bias (m) NOGAPS–QSCAT forcing	Rms error (m) NOGAPS forcing	Rms error (m) NOGAPS–QSCAT forcing	Correlation coef NOGAPS forcing	Correlation coef NOGAPS–QSCAT forcing
Atlantic	**0.04**	**0.40**	**−0.32**	**−0.14**	**0.58**	**0.51**	**0.92**	**0.92**
	0.04	0.06	−0.01	0.00	0.09	0.08	0.44	0.55
	0.06	0.08	−0.17	−0.12	0.39	0.35	0.71	0.73
	0.08	0.12	−0.41	−0.26	0.67	0.60	0.88	0.86
	0.12	0.40	−0.07	0.04	0.37	0.34	0.90	0.90
Pacific	**0.04**	**0.40**	**0.20**	**0.24**	**0.59**	**0.55**	**0.94**	**0.96**
	0.04	0.06	−0.22	−0.13	0.55	0.40	0.81	0.90
	0.06	0.08	0.20	0.25	0.54	0.52	0.90	0.93
	0.08	0.12	0.20	0.18	0.45	0.42	0.93	0.94
	0.12	0.40	0.05	0.06	0.28	0.26	0.93	0.94

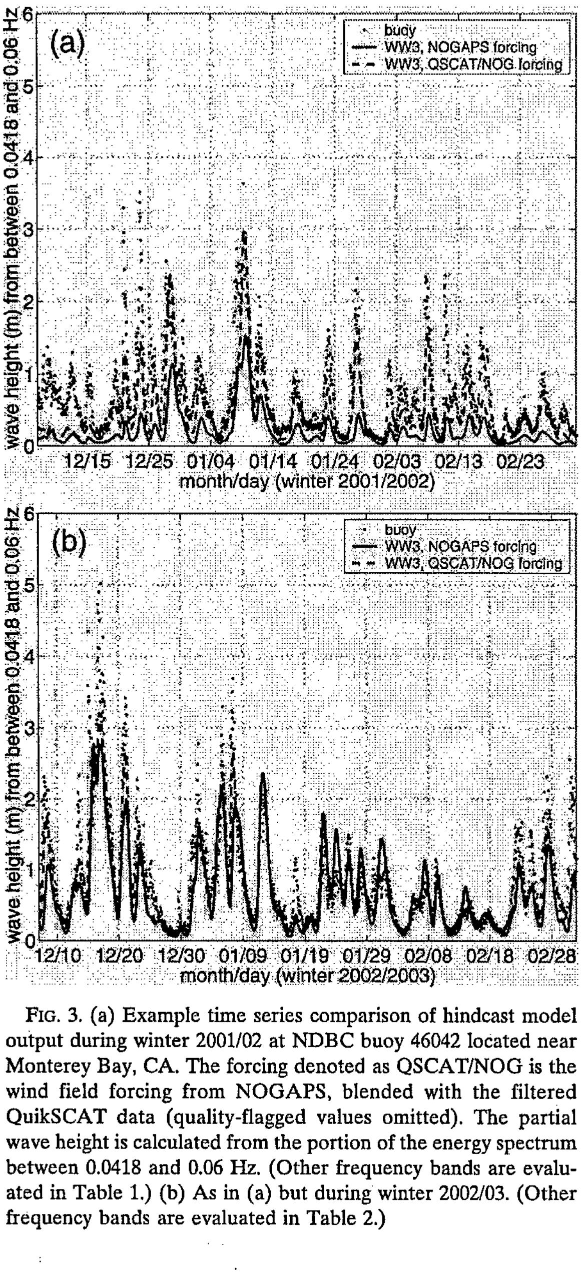

FIG. 3. (a) Example time series comparison of hindcast model output during winter 2001/02 at NDBC buoy 46042 located near Monterey Bay, CA. The forcing denoted as QSCAT/NOG is the wind field forcing from NOGAPS, blended with the filtered QuikSCAT data (quality-flagged values omitted). The partial wave height is calculated from the portion of the energy spectrum between 0.0418 and 0.06 Hz. (Other frequency bands are evaluated in Table 1.) (b) As in (a) but during winter 2002/03. (Other frequency bands are evaluated in Table 2.)

test 2a (in other words, bias from external error overwhelms any bias from the internal errors). To detect positive bias attributable to the wave model, we must rely exclusively on the winter 2002/03 hindcast.

b. Winter of 2002/03

During the winter of 2002/03 (Table 2), there is a negative bias in total energy of the models in the northwest Atlantic and a positive bias in the northeast Pacific. Since the QNF model tends to be more energetic than the NF model, the magnitude of bias is significantly lower than that of the NF model in the northwest Atlantic locations. At the northeast Pacific locations, the magnitude of bias in the QNF model is moderately higher than that of the NF model. The magnitude of the bias in total energy of the QNF model increased from the 2001/02 hindcast to the 2002/03 hindcast; this is likely due to a general increase in energy in the background wind vectors (i.e., NOGAPS analyses).

At the northwest Atlantic buoys, the bias is negative at all frequency bands. At the northeast Pacific buoys, the bias is negative in the band below 0.06 Hz and positive in the band above 0.06 Hz, suggesting an improper distribution of energy across frequencies. This trend is very different from that noticed in the northwest Atlantic, suggesting that the behavior is peculiar to the long fetch–duration situations typical of the northeast Pacific, or other climatological differences (e.g., prevalence of mixed sea states). In any event, this may be a problem that can be alleviated via tuning of source/sink terms (as in Tolman 2002d), but considerable further study would be required.

The correlation coefficient is higher for total wave height than for specific frequency bands; this probably reflects a tendency for errors at different frequencies to partially counteract each other.

The rms error of the QNF model for total wave height is slightly lower than that of the NF model. Thus, with the improvements made to NOGAPS during 2002, there is now only a slight advantage to supplementing the NOGAPS fields with QuikSCAT data in this manner.

Error metrics for the QNF wave hindcast are mostly better than those of the NF hindcast. This is despite the modestly better accuracy of the NOGAPS winds determined by Rogers et al. (2004) [see section 4c(2)].

In cases where the magnitude of the bias is nontrivial (say, greater than 7 cm), the sign of the bias of the NF model is identical to that of the QNF model. This is unexpected, since our direct validation of the wind fields suggests a bias of opposite sign. This clearly points to the conclusion that *bias in the 2002/03 wind fields is not a primary cause of bias in the wave hindcast results*. Using test 2a described in section 4a, we can say that some bias is probably attributable to the wave model itself, and from test 3a, it is (more specifically) likely associated with the deepwater source/sink term. Specifically, that bias shows the following:

- in the Atlantic, a negative bias in total wave energy;
- in the Atlantic, a negative bias at 0.06–0.12 Hz;
- in the Pacific, a negative bias at 0.04–0.06 Hz;

- in the Pacific, a positive bias in total wave energy; and
- in the Pacific, a positive bias at 0.06–0.40 Hz.

6. Discussion

As was discussed above, we chose not to apply test 1a in this study, because of the difficulty in defining and proving "small bias" in the wind fields. The concern is that (a) the accuracy of the ground truth itself cannot be absolutely proven (particularly a problem for higher wind speeds) and (b) the ground truth is geographically sparse. We believe, however, that the conclusions about the sign of bias of the wind fields (from Rogers et al. 2004) are probable enough to utilize in wave model evaluation.

Above, we drew the conclusion that bias in the 2002/03 wind fields is not a primary cause of bias in the wave hindcast results. Due to the inherent difficulty (or impossibility even) of systematically separating various sources of error, it is not possible to extend this conclusion to all locations and seasons.

It is worth stressing that given the necessary reliance on approximations in today's state-of-the-art wave models, it may be especially difficult for these models to have "universal" tuning. In particular, tuning for applications at one scale may inevitably degrade performance at another scale. For example, tuning to short-fetch empirical growth curves probably will not produce a skillful global model. Similarly, a model well tuned for basin-scale modeling may perform less well in subregional-scale applications. The question of generalized tuning is also discussed in Tolman (2002c). The tuning of models for various fetch–duration conditions is not explored or discussed in detail in this paper, though comparison of the fetch–duration relations of WAM4 and WW3 is presented in Rogers (2002). This comparison provides useful insight into the tuning of these models and how they relate to the Moskowitz (1964) data—which was the cornerstone of early WAM development (Komen et al. 1984)—and it may serve as a useful method of guidance in future tuning.

Though we do not verify that performance of the operational global WAM model (run at NAVO to create forcing for subregional wave models) is also improved due to the improvements to NOGAPS, it is a safe assumption that this is the case, since the hindcasts performed by Rogers (2002) and Rogers and Wittmann (2002) suggest that, given reasonably accurate forcing, WAM and WW3 are similarly skillful in predicting low-frequency energy (e.g., from the portion of spectra below 0.08 Hz) of windsea and young swells.

Some of the conclusions herein (such as the effect of NOGAPS improvements) are intended as evaluations of FNMOC's operational global WW3 implementation. We do not perform direct evaluation of operational products in this study. However, it is believed that conclusions based on these hindcasts should directly apply to the operational global wave analyses because of the "sanity check" performed by Rogers (2002) to verify that hindcast results closely matched operational analyses. Relevance to operational forecasts is less clear, since there tends to be some "drift" in bias of wave forecasts associated with drift in the wind forcing bias. Further study would be required to evaluate this. Janssen (1998) derived an error model for the wave model forecasts in term of the errors in the wind speed forecast. It is shown that a large portion of the forecast errors can be explained in term of errors in the forecast winds. These discrepancies between meteorological nowcast skill and forecast skill do not, of course, change observations herein about the WW3 model.

As mentioned in section 3b, swell attenuation in today's wave models is not expected to be accurate. So, in swell-dominated environments, we can still expect significant bias associated with inaccurate swell attenuation. In our applications of test 3a, we avoid specifically mentioning the accuracy of the wave model's representation of swell attenuation by including it as part of the more general "deepwater physics" along with the traditional "generation stage" source/sink terms (wind input, whitecapping, and four wave nonlinear interactions). The reason for this is simple: much of the swell attenuation is expected to occur geographically near the generation region, early in the swell energy's life cycle, since younger swells are steeper; thus, it is difficult to distinguish swell attenuation from the other three source/sink terms.

Using such data-derived wind fields to force wave models has a utility for real-time modeling that is not immediately obvious: though measured winds obviously cannot be used to forecast windseas, they can be used in the forecasting of much of the ocean's swell energy. The pertinent variables that determine which swells can be forecasted with measured winds are the age of the swell, the temporal range of the forecast, and the rapidity with which the data can be delivered and processed in real time. During 2003, a real-time system was created and run on an NRL workstation that created such fields, which were then used to force a global wave model run on the same workstation. The QuikSCAT data were delivered by FNMOC's Satellite Data Team to NRL usually within 2–4 h after the time of the measurements. Of course, if a ±6 h window of data is used, that incurs an additional 6-h delay. Results from this real-time wave model were very similar to those from the hindcast wave model described in this paper,

which uses the more slowly delivered PODAAC L2B data, suggesting that the fast delivery product from FNMOC is of high quality. Thus, it is demonstrated that with the recent advancement in the accessibility and geographic coverage of remotely sensed wind vectors epitomized by the QuikSCAT mission, ocean modelers have greater means with regard to the forcing used by their models.

Geographic variability is evident in the QuikSCAT fields, which do not exist in the operational analyses (the latter tend to be very smooth). Since the operational analyses are provided at intervals of, for example, 3 h, the fields are presumed to be representative of those 3 h. It is reasonable to expect a 3-h average of the "true" wind field to be smooth; thus, the operational analyses seem reasonable. However, from the standpoint of a wave model, the 3-h mean is not the only relevant parameter: the variability is also important. The spatial irregularity issue is similar to the temporal irregularity issue of gustiness. Komen et al. (1994) provide an estimate of the increase in wind input associated with gustiness (with no change in mean wind speed). [More recent treatment of this subject can be found in Abdalla and Cavaleri (2002), Abdalla (2001), and Abdalla et al. (2003).] In terms of standard statistics, current analyzed surface winds at a global scale have improved significantly in recent years but the lack of variability at smaller scales can be expected to result in systematic underestimation in wave generation (P. Janssen 2004, personal communication). The inclusion of inherently more variable winds from QuikSCAT (as has been done in this study) can be expected to reduce the systematic negative bias in the wave model (simultaneous with the reduction of wave energy bias via reduction of bias in the winds). This is an important point, because it implies that assimilation of scatterometer data with a state-of-the-art variational method (which will tend to produce smooth fields) may only address part of the problem (bias in wind speeds). Modifying the wind fields as was done here brings the variability of the data into the wind forcing. There is room for improvement, however: the variability in the QuikSCAT–NOGAPS wind fields is affected by inherent averaging scales (data resolution and model resolution). Ideally, wind speed should be provided to the wave model along with statistics about variability, but this requires further research and development.

We pointed out how, in the winter 2001/02 hindcast, the rms error of the QNF wave hindcast is lower than that of the NF wave hindcast, even though in the wind field validation (Rogers et al. 2004, summarized above) it was found that the rms error of the NOGAPS fields is less than that of the QuikSCAT–NOGAPS fields. This is presumably due to the nature of the wave model as a type of low-pass filter on wind fields: random, local errors in the wind field will tend to have little impact on wave model results, whereas systematic bias in moderate and/or high winds speeds will have a dramatic negative impact (unless there is a systematic bias in the wave model to balance the effect). This is important, as it suggests that a careful calibration of wind fields to remove systematic bias at specific wind speed bins may lead to improved wave predictions (this presumes, of course, that the wave model is skillful enough to convert improved forcing into improved wave predictions).

The improvement to the global wind forcing during the second half of 2002 led to the most noteworthy and unambiguous improvement to the Navy's operational global WW3 implementation in recent history. This in itself is a fairly flattering appraisal of the model, showing the maturity in the state of the art. Further improvements are likely to yield modest improvements to bulk error statistics, though it is still probable that major improvements can be achieved in error statistics local to certain geographic locations or frequency ranges.

In the validation by Rogers et al. (2004) of the four wind fields used in hindcasts herein, the bias of each wind speed "bin" is weighted (essentially) according to the square of the wind speed. We would point out that this is somewhat conservative (a higher power might be used instead), since it does not consider the frequency of the generated energy. Higher wind speeds generate energy at lower frequencies, which tends to persist longer in the ocean, thus disproportionately affecting the wave climate.

7. Conclusions

Conclusions from this study are listed below. They correspond to the three questions posed in the introduction (section 1c).

In the 2002/03 hindcasts comparison performed herein, bias in the wind forcing appears to be only a secondary source of bias in the wave model. Thus, further improvements to the wind field bias will not necessarily lead to improvements in wave predictions. In these hindcasts, more accurate wind forcing does not lead to more accurate wave predictions.

Comparison of the results from the wind field validations (in Rogers et al. 2004) and wave validations (herein) suggests that the bias of winds at moderate and high winds speeds, as a metric, is much more important to the skill of wave predictions than is wind speed rms error.

The clearest weakness in these hindcasts is a ten-

dency to significantly overpredict energy at higher frequencies and underpredict energy at lower frequencies. The frequency at which the bias changes sign is clearly different in the two oceans. In the northeast Pacific, it occurs at 0.06 Hz. In the northwest Atlantic, it occurs at 0.12 Hz or higher, if at all: the negative bias is observed over most of the model's frequency range. Apparent error associated with the physics in WW3 suggests that the model (and thus future operational nowcasts/forecasts) can benefit from additional tuning—perhaps similar to that performed by Tolman (2002d)—or some other upgrade to the physics.

Acknowledgments. The authors thank the FNMOC Satellite Data Team, Dr. Keith Sashegyi (NRL Marine Meteorology Division), Dr. James M. Kaihatu (NRL Oceanography Division), and Dr. Hendrik Tolman (NCEP/SAIC) for providing varied discussions, information, and resources. Feedback and recommendations from anonymous reviewers are also appreciated. This work was funded by the Office of Naval Research Ocean Modeling Program and the Naval Research Laboratory Core Program under Program Element 62435N.

REFERENCES

Abdalla, S., 2001: Impact of wind gustiness and air density on modelling of wave generation: Implementation at ECMWF. *Proc. ECMWF Workshop on Ocean Wave Forecasting,* Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 147–153.

——, and L. Cavaleri, 2002: Effect of wind variability and variable air density on wave modeling. *J. Geophys. Res.,* **107,** 3080, doi:10.1029/2000JC000639.

——, P. Janssen, and J.-R. Bidlot, 2003: Use of satellite data and enhanced physics to improve wave prediction. *Proc. 28th Int. Conf. on Coastal Engineering,* Cardiff, Wales, United Kingdom, American Society of Civil Engineers, 87–96.

Bender, L. C., 1996: Modification of the physics and numerics in a third-generation ocean wave model. *J. Atmos. Oceanic Technol.,* **13,** 726–750.

Bidlot, J.-R., D. J. Holmes, P. A. Wittmann, R. Lalbeharry, and H. S. Chen, 2002: Intercomparison of the performance of operational ocean wave forecasting systems with buoy data. *Wea. Forecasting,* **17,** 287–310.

Booij, N., and L. H. Holthuijsen, 1987: Propagation of ocean waves in discrete spectral wave models. *J. Comput. Phys.,* **68,** 307–326.

Cardone, V. J., H. C. Graber, R. E. Jensen, S. Hasselmann, and M. J. Caruso, 1995: In search of the true surface wind field in SWADE IOP-1: Ocean wave modeling perspective. *Global Atmos. Ocean Syst.,* **3,** 107–150.

——, R. E. Jensen, D. T. Resio, V. R. Swail, and A. T. Cox, 1996: Evaluation of contemporary ocean wave models in rare extreme events: The "Halloween storm" of October 1991 and the "storm of the century" of March 1993. *J. Atmos. Oceanic Technol.,* **13,** 198–230.

Chin, T. M., R. F. Milliff, and W. G. Large, 1998: Basin scale, high-wavenumber sea surface fields from a multiresolution analysis of scatterometer data. *J. Atmos. Oceanic Technol.,* **15,** 741–763.

Clancy, R. M., J. E. Kaitala, and L. F. Zambresky, 1986: The Fleet Numerical Oceanography Center Global Spectral Ocean Wave Model. *Bull. Amer. Meteor. Soc.,* **67,** 498–512.

Ebuchi, N., H. C. Graber, and M. J. Caruso, 2002: Evaluation of wind vectors observed by QuikSCAT/SeaWinds using ocean buoy data. *J. Atmos. Oceanic Technol.,* **19,** 2049–2062.

Emanuel, K. A., and M. Zivkovic-Rothman, 1999: Development and evaluation of a convection scheme for use in climate models. *J. Atmos. Sci.,* **56,** 1766–1782.

Fu, L.-L., E. J. Christensen, and C. A. Yamarone, 1994: TOPEX/POSEIDON mission overview. *J. Geophys. Res.,* **99,** 24 369–24 381.

Günther, H., S. Hasselmann, and P. A. E. M. Janssen, 1992: The WAM model Cycle 4 (revised version). Dtsch. Klim. Rechenzentrum Tech. Rep. 4, Hamburg, Germany, 101 pp. [Available from Institute for Coastal Research, GKSS Research Center, Max Planck Str. 1, D-21502 Geesthacht, Germany.]

Hogan, T. F., and T. E. Rosmond, 1991: The description of the U.S. Navy Operational Global Atmospheric Prediction System's spectral forecast models. *Mon. Wea. Rev.,* **119,** 1786–1815.

Janssen, P. A. E. M., 1998: On the error growth in wave models. ECMWF Research Memo. 249, ECMWF, Reading, United Kingdom, 12 pp.

——, B. Hansen, and J.-R. Bidlot, 1997: Verification of the ECMWF wave forecasting system against buoy and altimeter data. *Wea. Forecasting,* **12,** 763–784.

Jensen, R. E., P. A. Wittmann, and J. D. Dykes, 2002: Global and regional wave modeling activities. *Oceanography,* **15,** 57–66.

Khandekar, M. L., and R. Lalbeharry, 1996: An evaluation of Environment Canada's operational ocean wave model based on moored buoy data. *Wea. Forecasting,* **11,** 137–152.

Komen, G. J., S. Hasselmann, and K. Hasselmann, 1984: On the existence of a fully developed wind-sea spectrum. *J. Phys. Oceanogr.,* **14,** 1271–1285.

——, L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, and P. A. E. M. Janssen, 1994: *Dynamics and Modelling of Ocean Waves.* Cambridge University Press, 532 pp.

Moskowitz, L., 1964: Estimates of the power spectrums for fully developed seas for wind speeds of 20 to 40 knots. *J. Geophys. Res.,* **69,** 5161–5179.

Pierson, W. J., and L. Moskowitz, 1964: A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *J. Geophys. Res.,* **69,** 5181–5190.

PODAAC, 2001: QuikSCAT science data product user's manual. Jet Propulsion Laboratory Tech. Doc., Physical Oceanography Distributed Active Archive Center, 86 pp.

Rogers, W. E., 2002: The U.S. Navy's global wind-wave models: An investigation into sources of error in low frequency energy predictions. NRL Formal Rep. 7320-02-10035, 63 pp. [Available online at http://torpedo.nrl.Navy.mil/tu/ps.]

——, and P. A. Wittmann, 2002: Quantifying the role of wind field accuracy in the U.S. Navy's global ocean wave nowcast/forecast system. NRL Memo. Rep. 7320-02-8290, 26 pp. [Available online at http://torpedo.nrl.Navy.mil/tu/ps.]

——, J. M. Kaihatu, H. A. H. Petit, N. Booij, and L. H. Holthuijsen, 2002: Diffusion reduction in an arbitrary scale third generation wind wave model. *Ocean Eng.,* **29,** 1357–1390.

——, P.A. Wittmann, D.W. Wang, M. Clancy, and L. Hsu, 2004:

Evaluations of global wind prediction at Fleet Numerical Meteorology and Oceanography Center (from the perspective of a wave modeler). NRL Memo. Rep. 7320-04-8823, 15 pp. [Available online at http://torpedo.nrl.navy.mil/tu/ps.]

Rosmond, T. E., J. Teixeira, M. Peng, T. F. Hogan, and R. Pauley, 2002: Navy Operational Global Atmospheric Prediction System (NOGAPS): Forcing for ocean models. *Oceanography*, **15**, 99–108.

Teixeira, J., and T. F. Hogan, 2001: A new boundary layer cloud scheme in NOGAPS. Tech. Rep. NRL/MR/7532-01-7243, Marine Meteorology Division, Naval Research Laboratory, Monterey, CA, 38 pp.

——, and ——, 2002: Boundary layer clouds in a global atmospheric model: Simple cloud cover parameterizations. *J. Climate*, **15**, 1261–1276.

Tolman, H. L., 1991: A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents. *J. Phys. Oceanogr.*, **21**, 782–797.

——, 1999: Validation of NCEP's ocean winds for the use in wind wave models. *Global Atmos. Ocean Syst.*, **6**, 243–268.

——, 2002a: User manual and system documentation of WAVEWATCH-III version 2.22. NCEP Tech. Note, 133 pp. [Available online at http://polar.ncep.noaa.gov/waves/references.html.]

——, 2002b: Alleviating the garden sprinkler effect in wind wave models. *Ocean Modelling*, **4**, 269–289.

——, 2002c: Validation of WAVEWATCH III version 1.15 for a global domain. NCEP Tech. Note, 33 pp. [Available online at http://polar.ncep.noaa.gov/waves/references.html.]

——, 2002d: Testing of WAVEWATCH III version 2.22 in NCEP's NWW3 ocean wave model suite. NCEP Tech. Note, 99 pp. [Available online at http://polar.ncep.noaa.gov/waves/references.html.]

——, 2003: Treatment of unresolved islands and ice in wind wave models. *Ocean Modelling*, **5**, 219–231.

——, and D. Chalikov, 1996: Source terms in a third-generation wind wave model. *J. Phys. Oceanogr.*, **26**, 2497–2518.

——, B. Balasubramaniyan, L. D. Burroughs, D. V. Chalikov, Y. Y. Chao, H. S. Chen, and V. M. Gerald, 2002: Development and implementation of wind-generated ocean surface wave models at NCEP. *Wea. Forecasting*, **17**, 311–333.

WAMDI Group, 1988: The WAM model—A third generation ocean wave prediction model. *J. Phys. Oceanogr.*, **18**, 1775–1810.

Wittmann, P. A., 2001: Implementation of WAVEWATCH-III at Fleet Numerical Meteorological and Oceanography Center. *Proc. MTS/IEEE Conf. and Exposition: An Ocean Odyssey*, Honolulu, HI, Marine Technology Society and IEEE, 1474–1479.

——, and R. M. Clancy, 1993: Implementation and validation of a global third-generation wave model at Fleet Numerical Meteorological and Oceanography Center. *Proc. Second Int. Symp. on Ocean Wave Measurement and Analysis*, New Orleans, LA, Army Corps of Engineers, 406–419.

——, and W. C. O'Reilly, 1998: WAM validation of Pacific swell. *Proc. Fifth Int. Workshop on Wave Hindcasting and Forecasting*, Melbourne, FL, Atmospheric Environment Service/Environment Canada, 83–87.

——, R. M. Clancy, and T. Mettlach, 1995: Operational wave forecasting at Fleet Numerical Meteorology and Oceanography Center. *Proc. Fourth Int. Workshop on Wave Hindcasting and Forecasting*, Banff, AB, Canada, Atmospheric Environment Service/Environment Canada, 335–342.